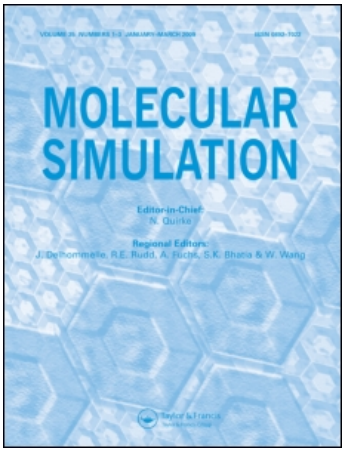

Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

Global optimisation by replica exchange with scaled hybrid Hamiltonians

Weixin Xu[a]; Ye Yang[a]; Yuguang Mu[a]; Lars Nordenskiöld[a]
[a] School of Biological Sciences, Nanyang Technological University, Singapore, Republic of Singapore

To cite this Article Xu, Weixin , Yang, Ye , Mu, Yuguang and Nordenskiöld, Lars(2008) 'Global optimisation by replica exchange with scaled hybrid Hamiltonians', Molecular Simulation, 34: 6, 575 — 590
To link to this Article: DOI: 10.1080/08927020801947020
URL: http://dx.doi.org/10.1080/08927020801947020

Molecular Simulation
Vol. 34, No. 6, May 2008, 575–590

Taylor & Francis
Taylor & Francis Group

Global optimisation by replica exchange with scaled hybrid Hamiltonians

Weixin Xu, Ye Yang, Yuguang Mu* and Lars Nordenskiöld

School of Biological Sciences, Nanyang Technological University, Singapore, Republic of Singapore 637551

(*Received 28 November 2007; final version received 24 January 2008*)

A global optimisation scheme based on replica-exchange molecular dynamics simulation with scaled hybrid Hamiltonians is presented and applied to fold trpzip2 peptide from extended structures with explicit water model using only eight replicas. The algorithm is shown to be capable of reproducibly optimising the peptide to structures of root mean squared deviations less than 2.5 Å with respect to all heavy atoms of NMR models. Moreover, the large amount of structural data sampled in the optimisation process enables us to provide a possible folding mechanism. The transition state ensemble is characterised by a largely formed turn and a compact packing of tryptophan 2, 4 and 9. The tryptophan 2/11 pair is found to form at a very late stage of the folding process. The first (closest to the turn) and the fourth native backbone hydrogen bonds form earlier and a picture of strict zipping up of hydrogen bonds is not observed. It is demonstrated in the present study that this global optimisation method which integrates structure prediction with approximated conformational samplings may be of help to understand the folding puzzle.

Keywords: global optimisation; all-atom simulation; replica-exchange with scaled hybrid Hamiltonian; trpzip2 β-hairpin

1. Introduction

It is a highly challenging problem for theoretical biophysicists to understand the relationship between protein sequence and its structure, and how a polypeptide chain finds its way towards the native structure from initial random coil states. Exponential increase of number of possible configurations results from the combinatorial monomer structures with the length of a polypeptide chain preventing simple energy scoring after direct enumeration. Efficient global minimum search methods have to be employed to tackle this protein folding puzzle. Among others, the simulated annealing (SA) algorithm [1,2] probably provided the first generally applicable technique for global optimisation. In this approach, the state of the system is followed by simulation as the temperature is decreased slowly from a high value, in the hope that it will eventually come to rest at the global potential energy minimum. Its success depended often strongly on the choice of the cooling schedule. Evidently, optimisation techniques are useful only when interest is restricted to predicting the structure of proteins, as they do not allow one to calculate thermodynamic averages.

One more aggressive method, which can not only obtain the lowest energy conformation, but also obtain any thermodynamic quantity at any temperature, is generalised ensemble simulations [3,4]. The idea is to achieve a random walk on potential energy surface. As a result the system can easily jump out of local minima and sample a much broader phase space. There are three well-known approaches for carrying out generalised ensemble molecular dynamics (MD) simulations: multi-canonical algorithm [5], simulated tempering [6,7] and replica-exchange method (REM) or parallel tempering [8]. Unlike the other two methods REM employs standard Boltzmann weight factors and thus is more easily accomplished. In this algorithm, a number of independent simulations (replicas) of the same system are performed in parallel, but under different conditions. At given time intervals, the replicas are exchanged with a specific transition probability between the replica pair. In its original implementation [8,9], the condition to be varied and exchanged among replicas was the temperature which resulted in a random walk in temperature space. Recently, REM has been increasingly employed to study folding energy landscape [10,11]. However for large systems, such as proteins in an explicit solvent, REM will require a huge number of computational facilities because the number of replicas needed increases as the square root of the degree of freedom of the system [12]. Moreover, the efficiency in sampling by REM begins to be noticed by some groups [13–15]: e.g. to fold *ab initio* a decapeptide and to get a converged folding thermodynamics more than 400 ns simulation was needed by REM explicit water simulation [13]. Endeavours to overcome the inadequacies of REM have

*Corresponding author. Email: ygmu@ntu.edu.sg

ISSN 0892-7022 print/ISSN 1029-0435 online

DOI: 10.1080/08927020801947020
http://www.informaworld.com

been made by several groups [12,16–19]. Among others an interesting REM with hybrid Hamiltonians (REMhH) was suggested by Simmerling's group and by us independently [20,21]. In this scheme, the dynamics of the system in each replica is controlled by system Hamiltonian which takes into account the explicit water molecules. During the replica exchange the energies which determine the number of replicas are calculated by taking solvation effect implicitly through a continuum solvent model, surface modified Generalised Born model (GBSA) [22].

It is interesting to compare REM with SA. The REM can be considered a self-controlled SA machinery. During the REM simulation process, configuration in each replica can be heated up or cooled down, switched by the total potential energy based on Metropolis criterion. Usually, the total potential energy is a suitable variable to characterise structural changes of the system. However in a complex system, such as protein solvated in explicit water, more than 90% of the total potential energy comes from water–water interaction. Thus, the total potential energy is no longer a good reaction coordinate for describing protein folding dynamics. Efforts to perform global optimisation of protein structure in explicit water using REM are less efficient because of the poor correlation between total potential energy and structure of protein [21]. Recently, Periole and Mark also found in the simulations of a β-heptapeptide that replica exchanges are accepted primarily based on random fluctuations within the solvent and are not strongly correlated with the instantaneous peptide conformation. This raises questions as to the efficiency of temperature REM in larger systems, although temperature REM is much more efficient than conventional MD simulations for rapidly equilibrating systems [15]. The introduction of a continuum solvent model in REMhH for calculating only the exchange probability makes the self-controlled SA machinery work more efficiently with regard to the dynamics of protein because continuum solvent models, like GBSA, estimate the free energy of solvation based solely on coordinates of the protein atoms. That is, we think, another merit of REMhH, besides the fact that the number of replicas can be greatly reduced.

Due to the hybrid Hamiltonian nature of this method, REMhH may not often guarantee the canonical ensembles in a strictly quantitative way. So REMhH is better termed as an efficient global optimisation method instead of a modified sampling strategy. On the other hand, all the ensembles from REMhH are created under constant temperatures with explicit consideration of protein–water interactions. In this work, the GBSA energies which are used in replica-exchange probability calculation are uniformly scaled by a constant less than 1 to further increase the temperature gap between neighbouring replicas without influencing the correlation between the GBSA energies and protein structure [21] to reduce the number of replicas further. In the meantime a reasonably high exchange rate can be maintained. We term this method a replica-exchange molecular dynamics simulation with scaled hybrid Hamiltonians (REMshH).

To demonstrate the applicability of the REMshH method, we try to predict the native structure of a short peptide solely from its amino acid sequence and to probe its folding mechanisms. It is noted that, presently, our understanding of β-sheet structures lags behind that of α-helices mainly because of a shortage of stable model systems for β-sheet secondary structures in isolation [23]. Recently, there emerged a kind of β-hairpins, such as tryptophan zippers (trpzip), which is of small size, unusual stability and very favourable spectroscopic properties [24]. Understanding the relative contributions of turn, strand and sidechain interactions to β-hairpin folding kinetics and thermodynamics will improve our understanding of β-sheet structures and further provide glimpses into the folding study of large proteins. However, although many works [25–46] have been done, the folding mechanism of β-hairpins remains vague. Various viewpoints are attained according to experimental observations with different techniques or simulation results obtained with different kinds of sampling methods or force fields. Whether it folds by a hydrogen bond-centric zipper mechanism [27,28,31,32] or a hydrophobic core-centric mechanism [29,30,33–35] or just a blend of the hydrogen bond-centric and hydrophobic-core centric mechanism [37,47] is still under debate. Eaton and coworkers [27,31] suggested a zipper mechanism in which folding is initiated from the turn and propagates towards the termini by forming hydrogen bonds. The hydrophobic cluster forms relatively later although it also plays an important role. Bonvin and Van Gunsteren also found that the turn is formed first, followed by hydrogen bond formation closing the hairpin, and subsequent stabilisation by hydrophobic interactions during the folding of a 19-residue β-hairpin [48]. In contrast, Karplus and coworkers [30] suggested that the hydrophobic cluster forms early, which brings together the main chain in this region; then the sheet propagates outward. Pande and coworkers [29,33] argued that the establishment of interactions between hydrophobic core residues is the central event and most probably the rate-limiting step. In addition, some works have shown that both the hydrogen bonds and hydrophobic interactions play important roles in mediating the folding kinetics [26,36–39,41,47]. Recently, by comparing the folding and unfolding rates of trpzip series [49], it was found that mutations of the hydrophobic core lead to increased unfolding rates, whereas mutations of the turn or loop regions primarily

affect the folding rate, which suggests that the rate-limiting event in the β-hairpin folding process corresponds to the formation of the turn, whereas the hydrophobic cluster only increases the stability of a β-hairpin. However, based on dynamic NMR study of GB1 peptide mutants, a change at the chain termini was found to increase folding rate 1.9-fold, which is difficult to rationalise by 'zipper' folding pathways [42]. Therefore, in order to make a contribution to solve this puzzling problem extensive MD simulation studies aiming at *ab initio* folding of short β-hairpin peptide with a full atomic physical force field are presented here.

2. Model and methods

2.1 Materials and simulation method

The trpzip2 peptide has 12 amino acid residues with the sequence of SWTWENGKWTWK. The native structure (PDB 1LE1) is shown in Figure 1 (the red ribbon). The N and C-terminals were protected by ACE and NME, respectively. The GROMACS program suite [50] and the full atomic OPLS-AA force field [51] were used. The simulation was carried out at constant volume. The explicit water model, TIP3P, was used under periodic boundary conditions. The peptide was solvated in a cubic box containing 382 water molecules with box length 24 Å. This box was big enough to avoid interactions between the folded conformations of the peptide and its periodic image but sufficiently small to enable some interaction in the fully extended state. A Cl^- ion was added to neutralise the charge of the system. All bonds involving hydrogen atoms were constrained in length according to LINCS protocol [52]. This allowed use of an integration step of 0.002 ps in simulations. Non-bonded pair lists were updated every five integration steps. The protein and the water were separately coupled to an external heat bath

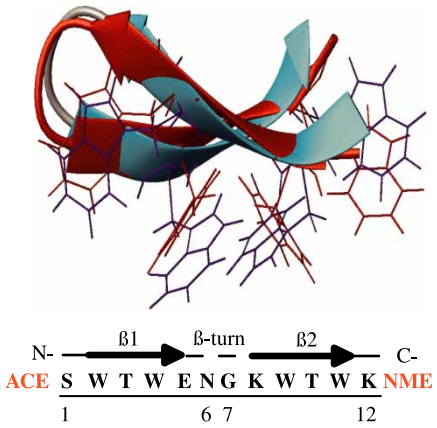

Figure 1. The ribbon structures of native (Model 1 of NMR structures, red) and folded (blue) trpzip2 β-hairpin. The four tryptophan residues of each structure are shown in neon.

with a relaxation time of 0.1 ps. Simulation data were recorded every 1 ps. The replica exchange was attempted every 2 ps. The GBSA energy was calculated by Tinker program [53]. GROMACS source code was modified to integrate with Tinker. Electrostatic interactions were calculated by using PME method with a cutoff of 10 Å, and the same cutoff was used in the calculation of van der Waals interactions. As mentioned in the Introduction, the GBSA energies were uniformly scaled by a constant value of 0.2. Thus, the system can be simulated using only eight replicas covering a temperature range from 260.5 to 700.8 K (260.5, 300.0, 345.6, 398.1, 458.6, 528.2, 608.4 and 700.8 K) with average exchange rate of 30%. The starting structures of the eight replicas were fully extended, from which 200-nsec trajectories were obtained. In order to check the repeatability of the simulation results three more simulations were performed with different initial velocities.

2.2 Analysis method

The weighted histogram analysis method (WHAM) is a general technique for combining data from multiple temperature-biasing simulations to obtain thermodynamic properties [54–57]. As an example, Gallicchio et al. [55] have shown that the WHAM (which they named as T-WHAM) can be successfully applied to the calculation of potentials of mean force by combining data from all the replicas of a replica-exchange simulation rather than using data only from the temperature of interest. Here, the total energy of the simulation system (with explicit water) was used in WHAM calculations. Two-dimensional WHAM as well as one-dimensional WHAM was employed here. Quantities calculated to describe the folding dynamics are the fraction of native contacts, Q, root mean squared deviations (RMSD) with respect to the NMR structure, RMSD for tryptophan pairs TRP2/TRP9 and TRP4/TRP9 and RMSD for the turn residues, the first and second principal components, PC1 and PC2, obtained from the dihedral principal component analysis (dPCA) [60] on both peptide backbone and backbone plus side chain dihedral angles, number of native backbone hydrogen bonds, radius of gyration for the whole peptide, Rg, radius of gyration for the hydrophobic core Rgcore (i.e. the four tryptophan residues: TRP2, TRP4, TRP9 and TRP11, see Figure 1). The radius of gyration, RMSD and Q, are calculated using all heavy atoms. The contacts are atom-based, defined as those atom-pairs within 5.0 Å of each other when the two amino acid residues are separated by four or more amino acids. Here, the common 141 contacts of 20 NMR structures in the PDB file (1LE1) are considered as the total number of native contacts. We found that the choice of threshold of 5.0 Å is not critical and that tuning this

value within the range 4.5–5.5 Å does not change the results qualitatively. A hydrogen bond is counted if the distance between two heavy atoms (N and O in this case) is less than 3.5 Å and the angle N—H···O is larger than 120°. There are five native backbone hydrogen bonds in the trpzip2 hairpin. To characterise the structures of the species identified from the free energy surface, we cluster the corresponding ensembles using the algorithm proposed by Snow et al. [61]. The conformations are clustered by their structural similarity, measured by pairwise RMSD. The related RMSD cutoff is set as 2.0 Å. All the statistical analysis reported here was performed on the last 150 ns of trajectories, while the first 50 ns were regarded as an equilibration phase. Free energy maps were checked for statistical convergence by generating them over increasingly longer simulation periods, 20, 30 ns, etc. No noticeable deviations were observed for free energy maps yielded at simulation intervals of 40 and 50 ns. Very recently, Baumketner and Shea also applied this method to determine the statistical convergence [62].

Like standard REM, REMshH also samples a wide conformational space at the cost of eliminating the information of dynamics. The trajectory in each replica is discontinuous due to the exchange of conformations between neighbouring temperatures. Here, we name the trajectory with constant temperature the 'ensemble trajectory'. However, if we tail after one conformation following the exchange information, a continuous 'folding trajectory' can be obtained. Such a trajectory is called a 'replica trajectory'. Recently, such a kind of trajectory has been used by Sanbonmatsu and Garcia [63] and Zhang et al. [47] in studying the folding transition of small proteins. All the structures are displayed with the program MOLMOL [64].

3. Results and discussion

3.1 Successful replica folding trajectory

Trpzip2 peptide is observed to fold from fully extended conformations into its native structure. As shown in Figure 2, evolutions of some quantities, such as temperature T, PC1, RMSD, Rg and Q as a function of time, demonstrate a successful replica folding trajectory. The simulation temperature plotted is obtained by averaging every 20 data points of the origin data. After ~31.2 ns, the simulation temperatures for this conformation firmly stay in low temperature regions. This is a good demonstration of how REMshH functions as a global optimisation scheme: it will bring the most stabilised structures down into low temperature replicas. After 31.2 ns, PC1, RMSD, Rg and Q are equilibrated at around −3.0, 2.0, 6.47 and 0.7 Å, respectively. The atom-based Q even reaches about 0.85 after 188.6 ns. The value of Q of the best folded structure can be as high as 0.94. It should be pointed out that the definition of atom-based

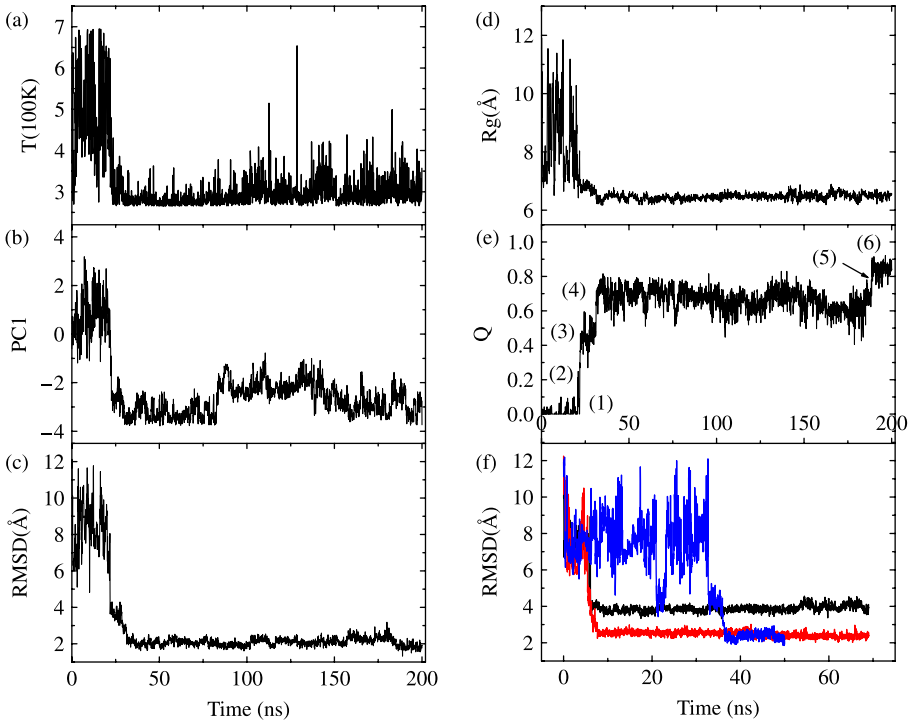

Figure 2. The time evolution of simulation temperature (a), the first principal component, PC1, from dPCA (b), RMSD (c), radius of gyration, Rg (d) and fraction of native contact, Q (e), and the best folded replica trajectories derived from three additional independent simulations (f).

Q is much stricter than the previously used residue-based definition of Q. A structure with $Q = 0.7$ is already much native-like (with RMSD for all heavy atoms less than 2 Å). This is very interesting because it is the first time that a peptide like trpzip2 is globally optimised into a native structure in the presence of explicit water molecules. One of the folded structures which was randomly chosen from the biggest cluster of the folded ensemble is shown in Figure 1, the blue trpzip2 β-hairpin. Its RMSD is of 1.172 Å with respect to the native NMR structure. Cross-stranded tryptophan rings pack intimately against one another, with fewer contacts between adjacent tryptophan pairs, similar to the NMR structure [24]. To test the repeatability of this optimisation process, three more simulations were performed with the same extended structures as initial structures but with different initial velocities. Figure 2(f) shows the best folded replica trajectories. One simulation was trapped in a conformation with RMSD around 4 Å. Two simulations succeeded in finding native structures with RMSD less than 2.5 Å. These successful optimisation simulations clearly claim the robustness of REMshH method in searching for the global minimum of peptides in the presence of explicit water molecules. Moreover, two other different peptide sequences have been *ab initio* folded into their native structures from initially extended conformation: one is GB1 hairpin with 16 amino acids and the other is mini-protein trpcage with 20 amino acids (data not shown).

To have a visual glimpse at this replica trajectory, some of the representative structures along the folding process are shown in Figure 3. Each structure belongs to a certain region which is marked in Figure 2(e). In the parenthesis below each structure, the value denotes the simulation time, temperature (T) and the values of Q, respectively. The folding picture emerges as follows: at the beginning, i.e. 0 ns, the peptide is fully extended with $Q = 0$ and $T = 528.6$ K. Two pairs of tryptophan residues (TRP4/9 and TRP2/11) are far away from each other. At 14.853 ns, T is changed to 608.4 K. The peptide collapses with $Q = 0.02$. The hydrophobic residues are non-specifically packed. At 21.881 ns, T is 458.6 K. The framework of the peptide looks somewhat native-like with $Q = 0.25$, but no secondary structures exist. TRP4 and TRP9 are close to each other while TRP11 is pointed to solvent. At 22.214 ns, the peptide is native-like with $Q = 0.45$, having two β strands and the turn largely formed. TRP4/9 is packed intimately and TRP11 is pointed to the other face of the peptide. At 31.189 ns, the structure is similar to that at 22.214 ns, except for the longer strands and a higher $Q = 0.56$. TRP11 begins to join the tryptophan cluster. At 45.047 ns, the peptide is much native-like with $Q = 0.75$. TRP4/9 and TRP2/11 are well packed. At 200 ns, the peptide is almost as the same as the native trpzip2 β-hairpin with $Q = 0.92$, RMSD = 1.7 Å. Therefore, Figure 3 shows a vivid folding picture of trpzip2 β-hairpin. Such a folding picture is obtained by a trajectory jumping between different temperatures. Due to the balance of forces modified with temperature [67], care should be taken in interpreting such a folding trajectory.

3.2 *Folding mechanisms of trpzip2*

One controversy about β-hairpin folding lies in the folding order between the turn and hydrophobic core. Herein, we try to get some clues from trpzip2 β-hairpin folding study. Various kinetic or thermodynamic 'phases', namely the molten globular state (MS), the transition state (TS), the intermediate state (IS) and the folded state (FS), are characterised. The 20 NMR structures are termed as native states (NS). As can be clearly seen from the replica folding trajectory shown in Figure 2(e), there are clusters of structures distinguished, based on the value of Q. So we define the 'kinetic' folding intermediate state (kIS)

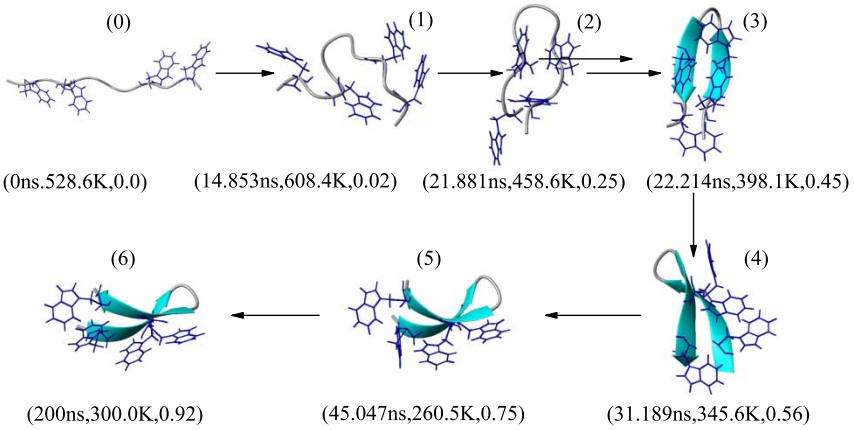

Figure 3. The representative structures extracted from the successfully folded replica trajectory.

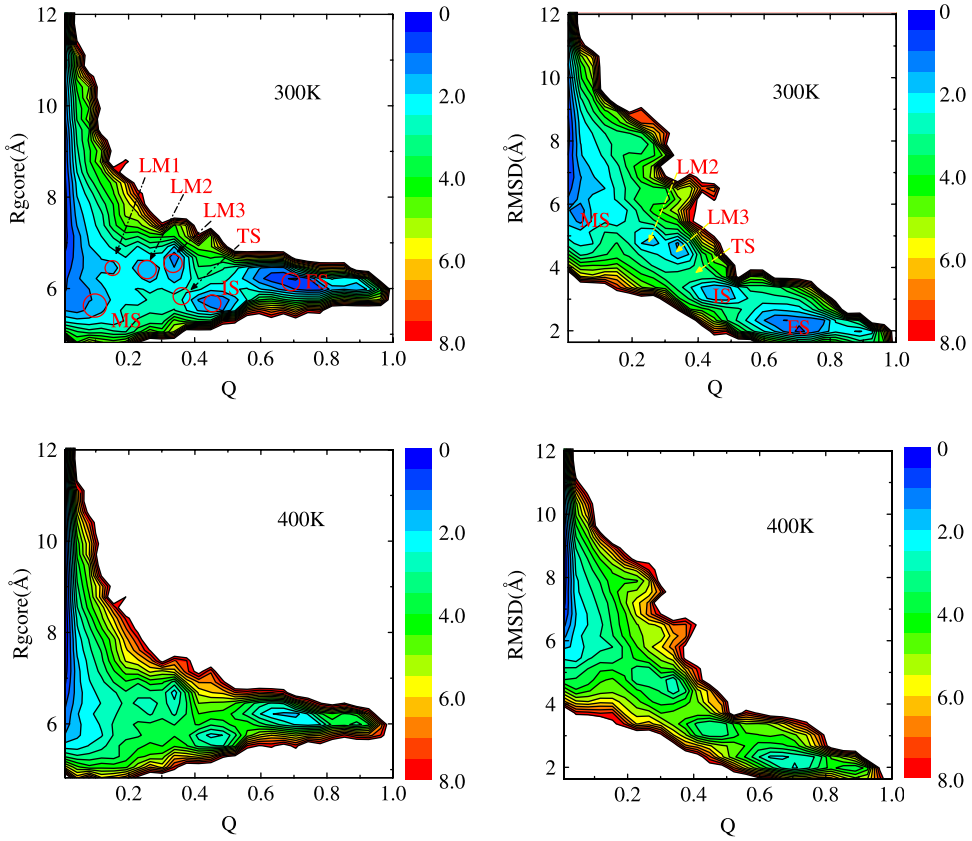

Figure 4. The contour plots of the free energy landscape at 300 K projected onto the reaction coordinates Q and Rgcore (a) and Q and RMSD (b) for a 200 ns simulation. LM1, LM2, LM3, MS, TS, IS and FS represent local minimum 1, local minimum 2, local minimum 3, molten state, transition state, intermediate state and folded state, respectively. Free energy landscape for the same 200 ns simulation at $T = 400$ K (c) and (d). All the plots are generated by two-dimensional WHAM.

at 22.21–31.09 ns with $0.36 \leq Q \leq 0.51$, the kinetic folded state (kFS) with $Q \geq 0.65$. Figure 4(a),(b) shows the free-energy landscapes at 300 K projected on a few reaction coordinates. Based on the free-energy surface, we define the local minimum 1 (LM1) at $0.13 \leq Q \leq 0.16$ and 6.29 Å $\leq$ Rgcore $\leq$ 6.35 Å, local minimum 2 (LM2) at $0.23 \leq Q \leq 0.28$, 4.6 Å $\leq$ RMSD $\leq$ 5.1 Å and 6.2 Å $\leq$ Rgcore $\leq$ 6.5 Å, local minimum 3 (LM3) at $0.32 \leq Q \leq 0.37$, 4.2 Å $\leq$ RMSD $\leq$ 4.7 Å and 6.4 Å $\leq$ Rgcore $\leq$ 6.8 Å, thermodynamic molten state (tMS) at $0.074 \leq Q \leq 0.12$ and 5.4 Å $\leq$ Rgcore $\leq$ 5.93 Å, the thermodynamic transition state (tTS) at $0.36 \leq Q \leq 0.41$, 3.6 Å $\leq$ RMSD $\leq$ 4.0 Å and 5.75 Å $\leq$ Rgcore $\leq$ 6.1 Å, the thermodynamic intermediate state (tIS) at $0.45 \leq Q \leq 0.50$, 3.0 Å $\leq$ RMSD $\leq$ 3.47 Å and 5.5 Å $\leq$ Rgcore $\leq$ 5.9 Å. The above Q, Rgcore and RMSD parameters are found to be insensitive to temperatures ranging from 300–450 K due to the fact that the locations of local minima on the free energy surface at 400 K (Figure 4(c),(d)) are nearly the same as those at 300 K (Figure 4(a),(b)) although distinct states have clearly different populations.

Then, we cluster the structures located in these regions and choose structures from the largest cluster as the representative ones. The corresponding features of some structural components and the global folding properties of trpzip2 are characterised and listed in Table 1. The first eight structures of each largest cluster are displayed in Figure 5. In Table 1, we first take a look at the kFS. Actually, the FS obtained from thermodynamic free-energy landscape is the same as the kinetic one. All the values, except the distance of the salt bridge between residue GLU5 and residue LYS12 (SB5/12) in the column of 'FS', are quite close to those of native values derived from the experimental NMR data [24]. The average Q, is equal to 0.7. and average RMSD = 2.0 ± 0.0 Å. On average 4.6 hydrogen bonds are formed. The corresponding folded structures are shown in Figure 5, labelled as 'kFS'. Obviously, both the backbone and four tryptophan residues are well assembled.

The turn is one of the important components of trpzip2. It is noted that both Rg and RMSD of turn at kinetic IS, and at thermodynamic TS and IS are extremely similar to that at FS. This means the turn of

Table 1. The statistical folding features of trpzip2.

		Kinetics		Thermodynamics					Statistics
Index		IS	FS	LM1	LM2	LM3	MS	TS	IS	NS
Turn	Rg	4.1 ± 0.2	4.0 ± 0.1	4.4 ± 0.2	4.0 ± 0.1	4.1 ± 0.1	3.7 ± 0.1	4.0 ± 0.1	4.1 ± 0.1	4.4 ± 0.2
	RMSD	1.8 ± 0.4	1.5 ± 0.3	2.4 ± 0.5	2.2 ± 0.2	1.4 ± 0.3	2.9 ± 0.1	1.6 ± 0.4	1.6 ± 0.4	0.0
Trp2/9	Rg	3.2 ± 0.2	3.5 ± 0.2	5.5 ± 0.2	5.5 ± 0.2	5.7 ± 0.3	4.7 ± 0.3	3.3 ± 0.2	3.5 ± 0.2	4.4 ± 0.1
	RMSD	3.9 ± 0.1	1.1 ± 0.3	2.9 ± 0.2	3.3 ± 0.1	3.4 ± 0.2	3.8 ± 0.1	3.9 ± 0.1	3.2 ± 0.2	0.0
Trp4/9	Rg	4.2 ± 0.3	3.7 ± 0.3	6.0 ± 0.3	7.3 ± 0.2	4.5 ± 0.2	4.2 ± 0.2	4.3 ± 0.3	4.7 ± 0.2	4.3 ± 0.1
	RMSD	1.4 ± 0.3	1.0 ± 0.3	4.1 ± 0.2	4.5 ± 0.2	2.3 ± 0.2	3.4 ± 0.1	1.4 ± 0.3	2.6 ± 0.2	0.0
Trp2/11	Rg	4.9 ± 0.3	4.0 ± 0.1	6.2 ± 0.4	4.5 ± 0.2	5.5 ± 0.4	3.4 ± 0.3	4.9 ± 0.2	4.6 ± 0.2	4.2 ± 0.1
	RMSD	2.7 ± 0.2	0.5 ± 0.1	3.7 ± 0.2	2.0 ± 0.1	4.0 ± 0.2	3.1 ± 0.4	2.8 ± 0.1	2.1 ± 0.2	0.0
HB		3.2 ± 0.8	4.6 ± 0.4	0.1 ± 0.0	1.8 ± 0.4	2.2 ± 0.4	0.0 ± 0.0	3.1 ± 0.9	3.8 ± 0.8	5.0 ± 0.0
SB5–8	Dist	9.2 ± 1.9	8.2 ± 1.7	10.5 ± 2.8	4.7 ± 1.5	7.5 ± 1.5	8.3 ± 0.5	7.9 ± 1.8	7.7 ± 1.8	9.1 ± 2.3
SB5–12	Dist	13.1 ± 1.9	3.8 ± 0.6	12.3 ± 1.9	9.3 ± 3.1	3.9 ± 0.7	7.3 ± 0.9	12.9 ± 1.8	4.5 ± 1.9	8.8 ± 2.9
Q		0.4 ± 0.0	0.7 ± 0.0	0.15 ± 0.08	0.25 ± 0.01	0.34 ± 0.01	0.1 ± 0.0	0.4 ± 0.0	0.5 ± 0.0	1.0 ± 0.0
Rg		6.9 ± 0.1	6.5 ± 0.1	7.1 ± 0.1	6.7 ± 0.1	6.6 ± 0.1	6.1 ± 0.1	7.0 ± 0.1	6.5 ± 0.1	6.9 ± 0.1
RMSD		3.6 ± 0.2	2.0 ± 0.1	5.3 ± 0.2	4.8 ± 0.1	4.4 ± 0.1	6.1 ± 0.1	3.8 ± 0.1	3.1 ± 0.1	0.0

Note: The list of various statistical and folding features for trpzip2. The kinetic and thermodynamic values are averaged over the biggest-cluster structures in corresponding regions. LM, MS, TS, IS, FS and NS represent local minimum, molten state, transition state, intermediate state, folded state and native state (20 NMR data), respectively.

Figure 5. The representative structures for states listed in Table 1. The characters k and t denote kinetic and thermodynamic, respectively.

trpzip2 forms in the early stage. The Rg of turn at MS is even smaller than those at FS and NS. Nevertheless, RMSD of turn at MS is much larger than that at FS, implying that turn is not shaped correctly at MS.

Another important structural component is the packing of four tryptophan resides (TRP2, TRP4, TRP9 and TRP11) which constitute the hydrophobic core. The Rg of TRP2/9 at both kinetic IS and thermodynamic TS and IS is smaller than that at FS, while at MS is significantly larger than that at FS. It means that TRP2 and TRP9 are packed close to each other at the TS and IS. The representative structures of kIS, tTS and tIS in Figure 5 clearly show that TRP2 and TRP9 are spatially close to each other. Similarly, all Rg of TRP4/9 pair at both kIS and tTS and tIS are close to values at FS. This demonstrates that TRP4 and TRP9 are packed together early in the folding process.

The situation is different for TRP2/TRP11 pair. All Rg at both kIS and tTS and tIS are much larger than that at FS. From the snapshots shown in Fig5, TRP2 and TRP11 are far away from each other at kIS, tTS and tIS. And especially, the indole ring of TRP11 is pointed to solvent from the other side of trpzip2 β-hairpin at these three states. This $Q \geq 0.55$ pattern lasts until the indole ring of TRP11 is switched to the same side where other indole rings of TRP2, TRP4 and TRP9 are located. Only afterwards is the TRP2/11 pair correctly packed. Based on RMSD of TRP2/9, TRP4/9 and TRP2/11 at each state listed in Table 1, the same conclusions can be drawn as mentioned above by observation of Rg values. In addition, from the snapshots at kIS, tTS and tIS shown in Figure 5, it is found that TRP2, TRP4 and TRP9 constitute a big and comparatively compact hydrophobic cluster and that TRP11 joins this hydrophobic cluster quite late, which indicates that the interaction of the more remote pair TRP2/TRP11 contributes less stabilisation than the TRP4/TRP9 pair. Very interestingly, our finding is very consistent with a recent extensive experimental study of β-hairpin peptides (trpzip analogs) which showed that deleting the outer tryptophan pair does not greatly affect the stability of the peptide [65].

Additionally, in our simulations, an obvious IS is detected. The IS we find both from kinetic analysis (kIS) and thermodynamics analysis (tIS) shares the same pattern: there is a heterogeneity of packing interaction among the four tryptophan residues. Most noted is tryptophan 11 which seems to experience fewer constraints from native topology.

The electrostatic interaction between charged residues is studied. GLU5 carries a charge of -1 and LYS8 and LYS12 have a charge of $+1$. A salt bridge (SB) is identified by the presence of at least one pair of atoms belonging to opposite charged groups, within a 3.5 Å distance [66]. Under this criteria, we find from 20 NMR structures an SB exists between GLU5 and LYS8. Here, we monitor the distances between CD atom belonging to GLU5 side chain, and NZ atom of LYS8 side chain (SB5/8), and the distance between the CD atom of GLU5 and the NZ atom of LYS12 (SB5/12) which are listed in Table 1. One can see that the distance of SB5/8 at FS, thermodynamic TS and IS is quite large. Similarly, the distance of SB5/12 at TS is also large. In the IS and FS the SB5/12 maintains a stable salt bridge. It seems that the salt bridge has certain thermodynamic stabilisation effects while contributing little to the folding kinetics. Interestingly, Thomas and Elcock's [67] simulation work did find that salt bridge interactions are extremely resilient to temperature increases and are uniquely suited to promoting protein stability at high temperatures.

3.3 *Unfolded ensemble*

3.3.1 *Local minima*

In this section, the unfolded ensemble is specially investigated. There are a few local minima in the unfolded ensemble on the free-energy landscape shown in Figure 4(a) with $Q \leq 0.4$. Three local minima, LM1, LM2 and LM3, are labelled on the surface and their properties are listed in Table 1. The representative structures are displayed in Figure 5. It is found that the unfolded species located within these local minima are not totally unstructured, but have significant native elements. For examples, structures of LM1 have a hairpin backbone. This hairpin has a reptation feature [41] with three backbone hydrogen bonds but none of them is native. The tryptophan residues are mismatched. The structures of LM3 possess a native-like turn and TRP4/9 pair. The RMSDs of turn and TRP4/9 are 1.4 and 2.3 Å, respectively (Table 1). The turn of structures in LM2 is disordered with RMSD around 2.4 Å. However, these structures have a native-like TRP2/11 pair and nearly two native HBs at the N/C terminal which may suggest that this β-hairpin folding could also be initiated far away from the turn. Clearly, our simulation samples a wide range of phase space of trizip2 and indicates the heterogeneity of the unfolded ensemble. Encouragingly, such heterogeneity was also reported in a combined experimental and computational study on the same trpzip2 peptide [68].

3.3.2 *Molten globular state*

It is also interesting to take a further in-depth look at the folding pattern of the MS which also belongs to the unfolded ensemble. Four of the largest clusters are extracted at MS as shown in Figure 6. All the structures in the four clusters are featured by the same average $Q = 0.1$, which implies MS is rather unstructured. The corresponding average radius of gyration of four classes of structures is within the range of 6.1 Å $\leq$ Rg $\leq$ 6.7 Å, which is smaller than that of the NMR native structures, 6.9 Å (Table 1). It is found that no hydrogen bond is formed in this state. Furthermore, the four hydrophobic tryptophan residues collapse in the four clusters in different ways. TRP2, TRP9 and TRP11 are packed together in cluster 1. In cluster 2, TRP2 and TRP9, TRP4 and TRP11 come together separately. TRP2 is close to TRP11 and TRP4 is glued to TRP9 in cluster 3. While in cluster 4, TRP2 and TRP4 are close, and TRP9 and TRP11 are packed together. There are reminiscent turn structures formed among residues 5–8 in cluster 1 and 2, while for cluster 3 and 4, the turn is strongly distorted. The average RMSD of turn for the four clusters is 3.0 Å with respect to the native turn indicating an unstructured turn. These observations suggest that the main folding driving force is originated from nonspecific hydrophobic interactions, which

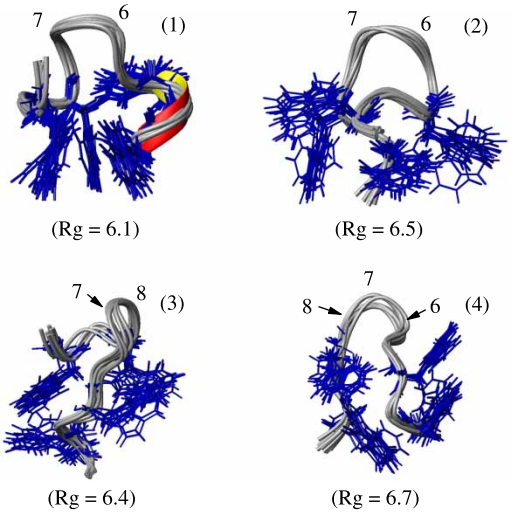

Figure 6. Four largest clusters of the MS. The numbers 5–8 refer to the residue numbering, which consists of the turn of trpzip2. All structures in these four clusters share a common characteristic that $Q = 0.1$ and no hydrogen bond is formed.

provides direct evidence for experimental findings that the very first step in β-hairpin folding corresponds to a process of hydrophobic collapse [42,46]. It is interesting to find that although it is a rather small peptide the folding features of trpzip2 are similar to those of large proteins.

3.4 Hydrogen bonds

The folding features of five native hydrogen bonds are studied in detail. Figure 7(a) is the formation probability of five hydrogen bonds as a function of time which is calculated based on the successfully folded replica trajectory. Around 5 ns, the first hydrogen bond (HB1) between oxygen atom O:GLU5 and hydrogen atom H:LYS8 begins to form. The formation probability is less than 0.1. At 20 ns the fourth hydrogen bond (HB4) between H:THR3 and O:THR10 starts to form. At about 21 ns, the third hydrogen bond (HB3) between O:THR3 and H:THR10 begins to form. Not until 22 ns does the second hydrogen bond (HB2) between H:GLU5 and O:LYS8 start to form. Finally, the fifth hydrogen bond (HB5) between O:SER1 and H:LYS12 starts to come into formation after 28 ns. Clearly, the order of formation of hydrogen bonds is: HB1 → HB4 → HB3 → HB2 → HB5. This is highly correlated with the folding mechanism of this β-hairpin. The turn and hydrophobic core TRP2/TRP4/TRP9 are formed almost simultaneously, which induces the early formation of HB1 and HB4. Evidently, the order of formation we find here does not support the strict zipping up model.

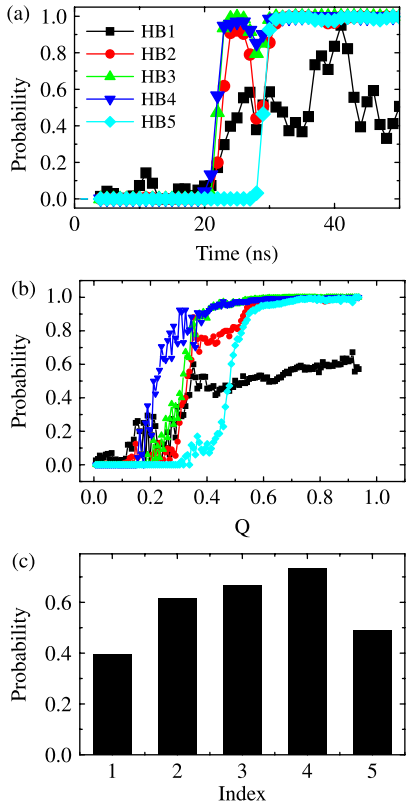

Figure 7. (a) The evolution of five hydrogen bonds as a function of time, which is averaged over every 2000 data points with 1000 data points overlapping. (b) The formation probability of each hydrogen bond as a function of Q. (c) The corresponding formation probability of each hydrogen bond. The above results are derived from the successfully folded replica trajectory.

As Q increases from 0 to 0.1, the formation probability of HB1 is around 0.04 and the probabilities for other four hydrogen bonds are ∼0.0 (Figure 7(b)). When Q increases to ∼0.18, the formation probability of HB1 reaches 0.28, significantly larger than those of other hydrogen bonds. The early formation of HB1 possibly indicates that HB1 plays an important role in stabilising the nascent turn at the early folding phase. However, HB1 hardly exists stably. Its formation probability is less than 0.7 even after all the other four hydrogen bonds are steadily formed. Figure 7(c) shows the formation probability of each hydrogen bond. HB3 and HB4 have the highest formation probabilities and HB1 and HB5 have the lowest formation probabilities. The populations of two hydrogen bonds close to the turn are lower than those centred on the strands. Both Dinner and Karplus's work [30] on peptide G β-hairpin and Zhang and Wang's study [47] on trpzip2 found the similar pattern of HB formation. The high formation probability of HB3 and HB4 may relate to the stabilisation role of the hydrophobic core which has been elegantly studied by Kelly group [69].

3.5 Reaction coordinates

Q and RMSD are well-known reaction coordinates. However, their application is mainly dependent on the definition of the native structure. It is desirable to have a 'natural' reaction coordinate which does not explicitly depend on the definition of NS but still can catch the main features of the dynamical process. Recently, a dPCA method based on a transformation of the peptide dihedral angles was applied to characterise the energy landscape of a tiny peptide [60]. Dihedral angles contain information mainly of local structure. It will be interesting to find out how well a quantity of optimised combination of local information (by dPCA) can be used to describe a global transition in peptide folding. Similar to canonical PCA results which work on cartesian coordinates the first few eigenvalues from dPCA are much larger than others. We find the first eigenvalue contributes more than 50% of the total variance. As a result, the first principal component, PC1, which is the projection of data of dihedral angles onto the first eigenvector, should provide a reasonable description of

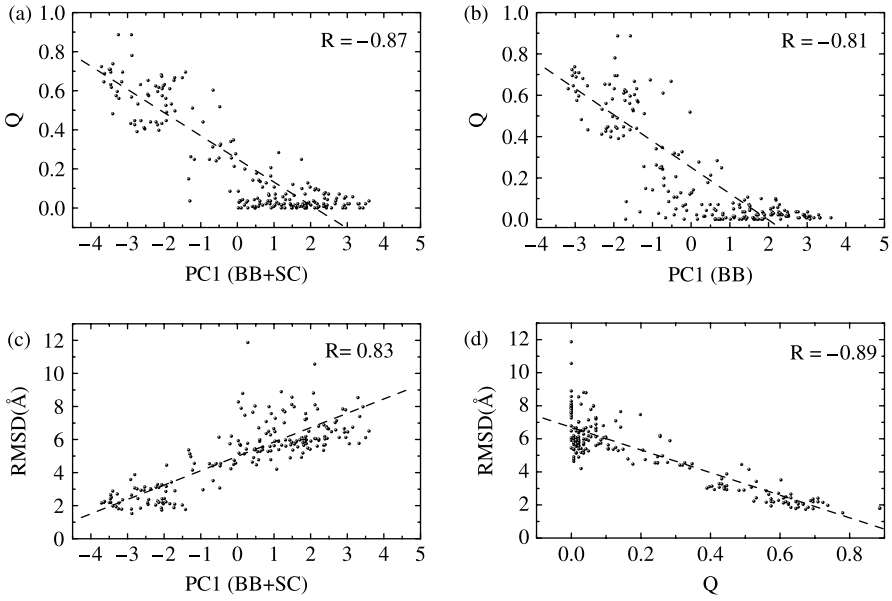

Figure 8. The correlations of reaction coordinates at 300 K between PC1 and Q (a, b), between PC1 and RMSD (c) and between Q and RMSD (d). Here, BB and BB + SC means that PC1 is based upon the dihedral angles of backbone atoms only, and of both backbone and side chain atoms, respectively. R denotes the correlation coefficient.

the dynamics of the system. The robustness of it as a reaction coordinate is verified by the degree of correlation between PC1 and two well-defined reaction coordinates, Q and RMSD. Two sets of dPCA are performed: one takes dihedral angles of both backbone and side chains atoms, the other considers dihedral angles of only backbone atoms. Good correlations between PC1 and Q and between PC1 and RMSD are obtained, -0.87 for PC1 and Q (Figure 8(a)), 0.83 for PC1 and RMSD (Figure 8(b)) based on both backbone and side chain atoms and -0.81 for PC1 and Q based on only backbone atoms (Figure 8(c)). For comparison, the correlation between the two well-defined reaction coordinates, Q and RMSD, is shown to be -0.89 (Figure 8(d)). The comparison clearly shows that PC1 is a good reaction coordinate comparable to Q and RMSD.

The one-dimensional distribution of PC1 (Figure 9(a)) shows two broad peaks with a border around PC1 $= -1.5$. The distribution of Q of the ensemble from the left peak (PC1 ≤ -1.5) (Figure 9(b)) demonstrates that this ensemble is mainly native-like, $Q \geq 0.4$. Thus, a two-state folding picture emerges, characterised by a local reaction coordinate, PC1. Then, a two-dimensional free energy surface using the first two principal

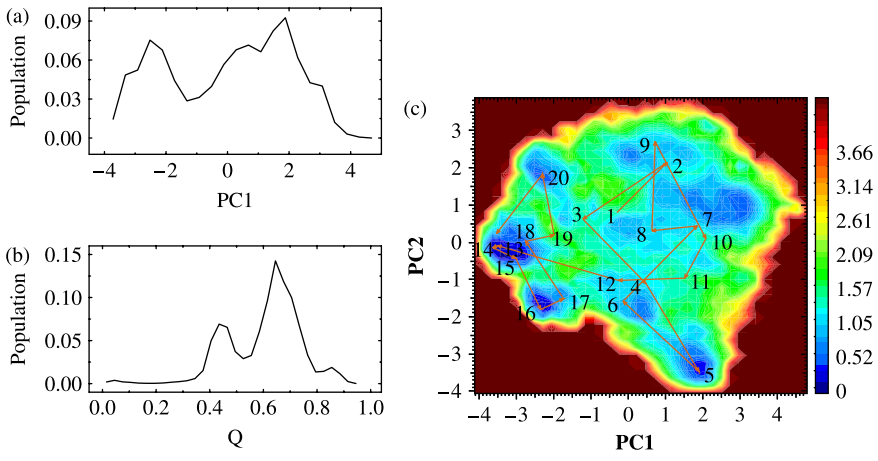

Figure 9. (a) The distribution of PC1. (b) The distribution of Q for those structures with PC1 ≤ -1.5. (c) The folding free-energy landscape contour plot at 300 K projected on the first two dPCA principal components, PC1 and PC2. The data are derived from the ensemble trajectory at 300 K.

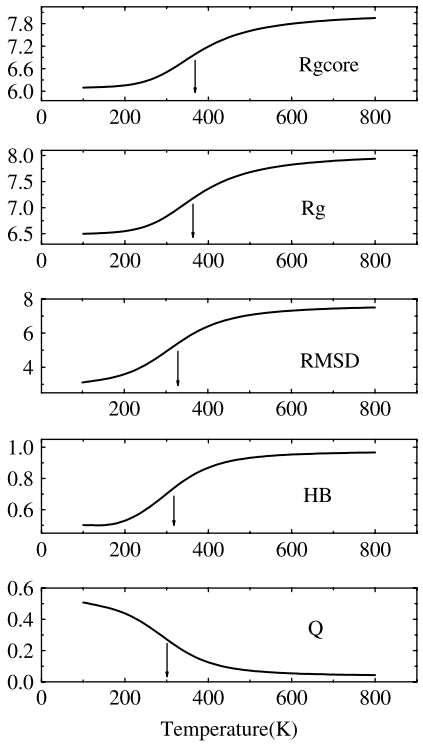

Figure 10. Melting curves of simulated observables. Top to bottom: radius of gyration of hydrophobic core ($T_m = 369$ K), radius of gyration ($T_m = 367$ K), RMSD ($T_m = 321$ K), average number of native backbone hydrogen bonds ($T_m = 308$ K) and fraction of native contacts ($T_m = 304$ K). The melting temperature T_m is marked by the vertical arrows which point to the midpoint of the simulated observable at low and high temperatures.

components, PC1 and PC2, as coordinates is constructed [59]. The data used are taken from the ensemble trajectory at 300 K. To clearly demonstrate the folding events, the data from the successful replica folding trajectory are projected onto the two-dimensional free energy surface as points linked by lines. The numbers on the figure label the simulation time for the replica trajectory as the same in Figure 2(b). The native basin with PC1 ≤ -1.5 has three well-resolved minima. The structural difference between the minima is due to different packing patterns of the four tryptophan residues, as mentioned earlier in the paper. Folding is progressed from the right half to the left half of the surface as indicated from the numbers. Therefore, our results prove that PC1, a quantity of combination of local information, is able to capture the principal conformational changes during the folding process. The success of dPCA in distinguishing the folded and unfolded states is due to the fact that in FSs the dihedral angles are well defined or are restricted by the native constraints. While in the unfolded states such constraints are missing.

Figure 10 shows several other commonly used reaction coordinates as a function of temperature

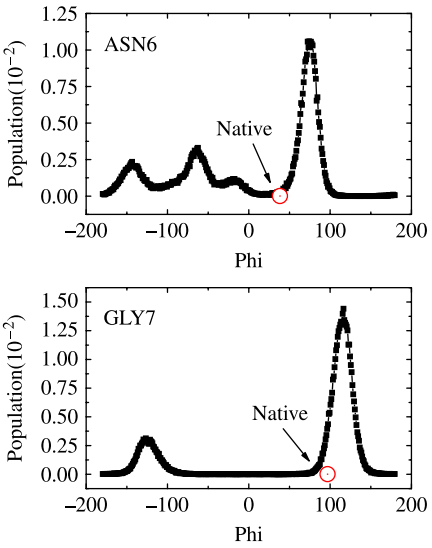

Figure 11. The distribution of angles of ASN6 and GLY7 on the turn of trpzip2 β-hairpin. The circle mark denotes the average value derived from 20 NMR native structures.

obtained by using WHAM. It is found that when monitored by Rgcore and Rg, the calculated T_m (around 368 K) is larger than the experimental value of 345 K [24,68]. However, when monitored by RMSD, HB or Q, the T_ms span a narrow temperature range from 304 to 321 K, which is smaller than the experimental T_m. Interestingly, both Yang et al. [68] and Zhang et al. [47] found that the T_m obtained by monitoring different observables differ significantly. This discrepancy may be attributed to the heterogeneity in the folding of trpzip2 hairpin [68].

3.6 Possible problems of the force field

Trpzip2 adopts a β-hairpin with a type I' β-turn centred at residues GLU5, ASN6, GLY7 and LYS8. The distribution of the backbone ϕ dihedral angles and GLY7 extracted from the successful folding trajectory together with the averaged value from 20 NMR structures are shown in Figure 11. The deviations of simulated values from NMR averages are evident. A study based on analysis of native protein structures observed that the average dihedral angle values (ϕ, ψ) for I' turn in β-structures are (54, 40) and (78, 4) degrees for residues at similar positions as *ASN*6 and GLY7 locates [70,71]. The average values of ϕ on *ASN*6 and GLY7 in our study are 74 and 115° which are 20 and 37° larger than the canonical values, respectively. To find the reason for the discrepancy the equilibrium structures of a model system, glycine dipeptide (Ace-GLY-Nme) solvated in water, was studied by MD simulation using OPLSAA force field. The average ϕ value is 118° in the related

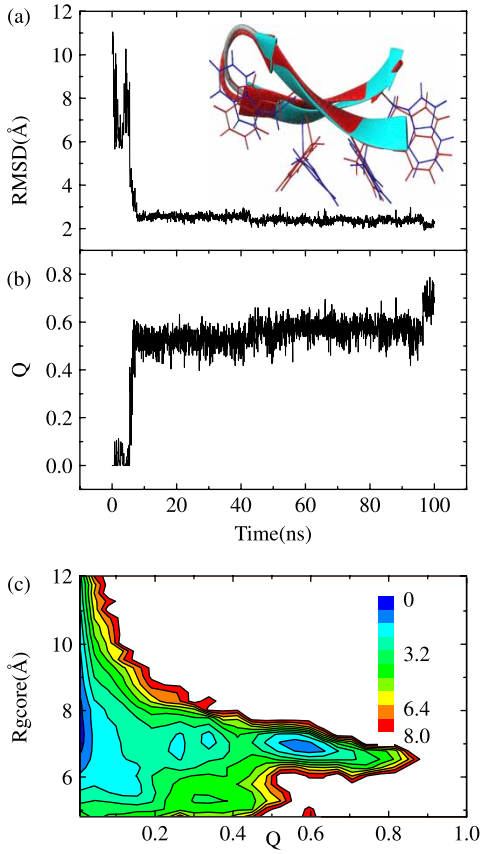

Figure 12. Results for unscaled REMhH MD simulations. The time evolution of RMSD (a) and the fraction of native contact, Q (b), both quantities are calculated based on one replica trajectory which is derived from 16 temperature ensembles by tracking one conformation according to exchange information. The inset in (a) is one representative folded structure (blue) compared with Model 1 of NMR structures (red). The contour plot of the free energy landscape at 300 K projected onto the reaction coordinates Q and Rgcore (c).

conformational ensemble. So the difference between simulated and experimental structures is mainly caused by the parameters defining the ϕ value. A modification on OPLSAA force field which only involves the backbone ϕ dihedral angles is suggested and further tested in *ab initio* folding several peptides including trpzip2. Encouraging results are obtained. Details will be presented in our further studies.

4. Discussion and conclusions

The successful folding of trpzip2 peptide in 30 ns from extended structures is attributed to several factors. Firstly, the optimisation algorithm based on REMshH plays an important role. It is a self-controlled repetitive SA process. The cooling/heating scheme is based on the implicit solvation energy of peptide, surface generalised Born solvation energy. It is found that the OPLS-AA force field combined with GBSA is able to identify correctly the native state of several large proteins [72]. Employing the same force field Ulmschneider and Jorgensen have found the global minimum of trpzip2 peptide which agrees closely with the NMR structures by Monte Carlo sampling with concerted rotation [73]. On the other hand, a replica-exchange molecular dynamics simulation study using OPLS-AA combined with GBSA indicates a nonnative global minimum for a β-hairpin peptide, C-terminal β-hairpin of protein G [74]. An overstabilised salt-bridge effect between charged residues is found to be responsible for this behaviour in the implicit solvent model. With REMshH scheme applied in this work the dynamics of peptide is governed by the Hamiltonian including explicit water molecules. The over-stabilisation effect could be greatly alleviated. Recent efforts to improve the GBSA model show prospective results [75,76]. It will be interesting to have a direct comparison of folding energy landscape of that with explicit water and that of improved implicit solvation models.

The number of replicas used in this study is only 8 which is even less than that used in completely implicit solvation model REMD simulations by Yang et al. [68] and by Zhou and Berne [74] where 28 and 18 replicas were used, respectively. The simulation temperature range covered is similar, from 260 to 700 K. The exchange acceptance ratios are around 30% in all cases (Figure 13). Evidently, the temperature gap between neighbouring replicas in our study is much larger (Figure 13). Maintaining similar acceptance ratio between replicas with much larger temperature gap violates detailed balance condition (DBC). The DBC requires that the exchange probability between two replicas at temperature T_i and T_j is determined by the value $e^{-\Delta}$, where $\Delta = ((1/k_B T_i) - (1/k_B T_j))(E_j - E_i)$, E_i and E_j are the potential energy of the two replicas, k_B is the Boltzmann constant. In such a way, canonical ensembles are guaranteed at all temperatures. The REMhH and REMshH method changes the way of calculating Δ. As a result, the canonical ensembles may not be maintained in a strictly quantitative way. However, the results shown here as well as previous ones [20,21] indicate that the violation of DBC in this way may not result in unphysical outcomes. How does it come that the DBC could be somehow loosened? We interpret this phenomenon in the following way. The DBC employed in the canonical REMD guarantees that the new configurations which are produced by rescaling velocities in the exchange processes conform well to corresponding canonical ensembles. In the REMhH or REMshH scheme, the new configurations created by exchange process may not be in accord with canonical ensembles because of violation of DBC. However, in practice, such configurations are not outputted as sampling data. In the current study we save

structural data every 1 ps. Thus, the new configurations will be equilibrated for 1 ps. Within the time scale of 1 ps, the water molecules have finished their libration motion [77] and peptide backbone atoms have completed their fast motion (around 100 fs) [78]. So the violation of DBC is somehow compensated by a enough equilibration through MD simulation under constant temperature.

In this study, a factor of 0.2 is used to uniformly scale the GBSA energies which are employed to determine the transition probability. This is our trial. Actually, both unscaled and scaled REMhH models are attempted in our simulation studies. For comparison, the REMhH simulation study without scaling of GBSA energies on the same peptide is presented. To cover the same temperature range and to maintain the same replica-exchange rates, 16 replicas are needed in the simulation. The simulation results are summarised in Figure 12. In one successfully folded replica trajectory, the peptide finds its native-like configurations from an extended conformation within 10 ns (Figure 12(a),(b)), which indicates a global optimisation process on the energy landscape. Also presented is the contour plot of the free-energy landscape at room temperature projected onto the reaction coordinates Q and radius of gyration of the hydrophobic core, R_{gcore}. Interestingly, the free-energy landscape constructed from the scaled simulation (Figure 4(a)) is quite similar to that obtained from the unscaled one (Figure 12(c)). There is the same number of local minima on the surfaces and their locations highly correspond to each other. However, the presence of some differences is also noted. The Q value of the native state from the scaled simulation is around 0.7 while that of the native state from the unscaled simulation is around 0.6. The local minimum which is located around $Q = 0.45$ and $R_{\mathrm{gcore}} = 5.8$Å from the unscaled ensemble is more shallow and broad than that of the scaled ensemble. In this sense, both unscaled and scaled REMhH models provide qualitatively similar simulation results.

The usage of a universal scale factor less than 1 for the exchange GBSA energies speeds up the conformation exchange between replicas with higher temperature gaps. On the other hand, the byproduct is mixing up conformations from different temperatures and resulting in conformation ensembles more deviated from canonical ones. So, as discussed in the Introduction section, such a method is an approximated sampling method. The GBSA energies (scaled or not) play as a scoring function to select native-like conformers out of configuration pools and to bring them down to lower temperature replica. Thus the free-energy landscape constructed by this method could only be interpreted qualitatively instead of quantitatively. Probably, due to the same reason, the reversible folding/unfolding events are not found in all trajectories. This may indicate the limitation of the current method whose exchange probability exclusively depends on GB energies. More rigorous methods which are able to reproduce the canonical ensembles and meanwhile to reduce the number of replicas are needed and such study is being undertaken by us.

In summary, an optimisation method using REMshH is presented and applied to fold trpzip2, a β-hairpin peptide from extended structures in explicit aqueous solution with OPLS-AA force field. The final optimised structures have all heavy atom-based RMSD smaller than 2.5 Å. With ample sampling collected during the optimisation process the folding mechanism is explored in details. The TS ensemble is mainly characterised by a largely formed turn

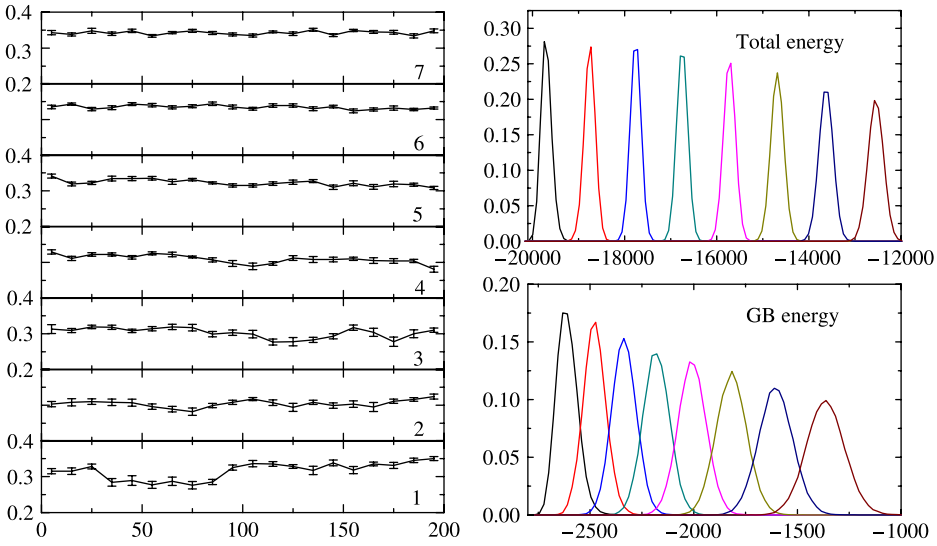

Figure 13. The exchange ratios as a function of time using windows of 10 ns (a). Distributions of the total energy and GB energy at each temperature (b).

and the compact packing of TRP2, TRP4 and TRP9 as well as 60% hydrogen bonds formed. Folding is cooperatively progressed and driven by the nonspecific collapse of hydrophobic cluster at very early stage. MS is thus observed and featured by the unstructured turn, no hydrogen bonds and nonspecific interactions of the four tryptophan residues. Among five native backbone hydrogen bonds, the 1st hydrogen bond closest to the turn forms earliest, followed by the 4th hydrogen bond. Consequently, no strict zipping up of hydrogen bonds is detected. A temporarily stable kinetic intermediate state is found to accumulate during the simulation featured by the alternative packing patterns of hydrophobic core. The TRP2/TRP11 pair is found to contribute minor stabilisation to the β-hairpin. The first principal component, PC1, from dihedral PCA analysis turns out to correlate well with well-known reaction coordinates, *Q* and RMSD.

Some results in our present study are based on the folding events observed in one successful folding trajectory. However, analysis of other successfully folded trajectories shown in Figure 2(f) provides similar results. Nevertheless, it should be kept in mind that alternative folding processes are also likely to occur given the complexity of the free energy landscape and the limitations of the current classical force field. The REMhH or REMshH method has intrinsic limitations rooted from the nature of hybrid Hamiltonians. The conformational bias induced by GBSA simulation protocols has been well documented [79] and several directions of improving implicit solvation model are being pursued by different groups [76,80,81] (Figure 13).

REMshH provides a robust framework to perform *ab initio* peptide structure prediction and folding mechanism exploration whose merits rely on the optimised searching of global minimum and efficient usage of computational facilities. With ever increasing computational power and advancing algorism more details about protein folding mechanism will be unraveled.

Acknowledgements

The support of research grants, SUG and URC(RG65/06), from Nanyang Technological University, and a Ministry of Education AcRF Tier 2 grant (T206B3210RS) is acknowledged by Y. M. and a Ministry of Education AcRF Tier 2 grant (T206B3207) by L. N. The simulations were performed on the supercomputer of Bioinformation Research Center in NTU and the Frankfurt Center of Scientific Computing, which are acknowledged by the generous allocation of CPU time.

References

[1] S. Kirkpatrick, C.D. Gelatt, and M.P. Vecchi, *Optimization by simulated annealing*, Science 220 (1983), pp. 671–679.
[2] D.J. Wales and H.A. Scheraga, *Global optimization of clusters, crystals, and biomolecules*, Science 285 (1999), pp. 1368–1372.
[3] U.H. E. Hansmann and Y. Okamoto, *New Monte Carlo algorithms for protein folding*, Curr. Opin. Struct. Biol. 9 (1999), pp. 177–183.
[4] A. Mitsutake, Y. Sugita, and Y. Okamoto, *Generalized-ensemble algorithms for molecular simulations of biopolymers*, Biopolymers 60 (2001), pp. 96–123.
[5] B.A. Berg and T. Neuhaus, *Multicanonical ensemble: A new approach to simulate first-order phase transitions*, Phys. Rev. Lett. 68 (1992), pp. 9–12.
[6] E. Marinari and G. Parisi, *Simulated tempering: A new Monte Carlo scheme*, Europhys. Lett. 19 (1992), pp. 451–458.
[7] A.P. Lyubartsev, A.A. Martsinovski, S.V. Shevkunov, and P.N. Verentsov-Velayaminov, *New approach to Monte Carlo calculation of the free energy: Method of expanded ensembles*, J. Chem. Phys. 96 (1992), pp. 1776–1783.
[8] U.H.E. Hansmann, *Parallel tempering algorithm for conformational studies of biological molecules*, Chem. Phys. Lett. 281 (1997), pp. 140–150.
[9] Y. Sugita and Y. Okamoto, *Replica-exchange molecular dynamics method for protein folding*, Chem. Phys. Lett. 314 (1999), pp. 141–151.
[10] R. Zhou, B.J. Berne, and R. Germain, *The free energy landscape for beta hairpin folding in explicit water*, Proc. Natl. Acad. Sci. USA 98 (2001), pp. 14931–14936.
[11] A.E. Garcia and J.N. Onuchic, *Folding a protein in a computer: An atomic description of the folding/unfolding of protein A*, Proc. Natl. Acad. Sci. USA 100 (2003), pp. 13898–13903.
[12] H. Fukunishi, O. Watanabe, and S. Takada, *On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction*, J. Chem. Phys. 116 (2002), pp. 9058–9067.
[13] M.M. Seibert, A. Partriksson, B. Hess, and D. van der Spoel, *Reproducible polypeptide folding and structure prediction using molecular dynamics simulations*, J. Mol. Biol. 354 (2005), pp. 173–183.
[14] D.M. Zucherman and E. Lyman, *A second look at canonical smapling of biomolecules using replica exchange simulation*, J. Chem. Theory. Comput. 2 (2006), pp. 1200–1202.
[15] X. Periole and A.E. Mark, *Convergence and sampling efficiency in replica exchange simulations of peptide folding in explicit water*, J. Chem. Phys. 126 (2007), pp. 014903–014911.
[16] P. Liu, B. Kim, R.A. Friesner, and B.J. Berne, *Replica exchange with solute tempering: A method for sampling biological systems in explicit water*, Proc. Natl. Acad. Sci. USA 102 (2005), pp. 13749–13754.
[17] T.Z. Lwin and R. Luo, *Overcoming entropic barrier with coupled sampling at dual resolutions*, J. Chem. Phys. 123 (2005), pp. 1949041–19490410.
[18] X. Cheng, G. Cui, V. Hornak, and C. Simmerling, *Modified replica exchange simulation methods for local structure refinement*, J. Phys. Chem. B 109 (2005), pp. 8220–8230.
[19] E. Lyman, F.M. Ytreberg, and D.M. Zuckerman, *Resolution exchange simulation*, Phys. Rev. Lett. 96 (2006), pp. 028104–028105.
[20] A. Okur, L. Wickstrom, M. Layten, R. Geney, K. Song, V. Hornak, and C. Simmerling, *Improved efficiency of replica exchange simulations through use of a hybrid explicit/implicit solvation model*, J. Chem. Theory. Comput. 2 (2006), pp. 420–433.
[21] Y. Mu and Y. Yang, *Modeling of water explicitly in the replica-exchange simulation method for protein folding*, NIC Series:http://www.fz-juelich.de/nic-series/volume34/mu.pdf. 34 (2006), pp. 119–224.
[22] D. Qiu, P.S. Shenkin, F.P. Hollinger, and W.C. Still, *The GB/SA continuum model for solvation: A fast analytical method for the calculation of approximate Born radii*, J. Phys. Chem. A 101 (1997), pp. 3005–3014.
[23] R.M. Hughes and M.L. Waters, *Model systems for β-hairpins and β-sheets*, Curr. Opin. Struct. Biol. 16 (2006), pp. 514–524.
[24] A.G. Cochran, N.J. Skelton, and M.A. Starovasnik, *Tryptophan zippers: Stable, monomeric β-hairpins*, Proc. Natl. Acad. Sci. USA 98 (2001), pp. 5578–5583.
[25] S. Gianni, N.R. Guydosh, F. Khan, T.D. Caldas, U. Mayor, G.W. White, M.L. DeMarco, V. Daggett, and A.R. Fersht,

Unifying features in protein-folding mechanisms, Proc. Natl. Acad. Sci. USA 100 (2003), pp. 13286–13291.
[26] M. Petrovich, A.L. Jonsson, N. Ferguson, V. Daggett, and A.R. Fersht, *phi-Analysis at the experimental limits: Mechanism of beta-hairpin formation*, J. Mol. Biol. 360 (2006), pp. 865–881.
[27] V. Munoz, E.R. Henry, J. Hofrichter, and W.A. Eaton, *A statistical model for β-hairpin kinetics*, Proc. Natl. Acad. Sci. USA 95 (1998), pp. 5872–5879.
[28] A. Kolinski, B. IIkowski, and J. Skolnick, *Dynamics and thermodynamics of β-hairpin assembly: Insights from various simulation techniques*, Biophys. J. 77 (1999), pp. 2942–2952.
[29] V.S. Pande and D.S. Rokhsar, *Molecular dynamics simulations of unfolding and refolding of a β-hairpin fragment of protein G*, Proc. Natl. Acad. Sci. USA 96 (1999), pp. 9062–9067.
[30] A.R. Dinner, T. Lazaridis, and M. Karplus, *Understanding β-hairpin formation*, Proc. Natl. Acad. Sci. USA 96 (1999), pp. 9037–9068.
[31] V. Munoz, P.A. Thompson, J. Hofrichter, and W.A. Eaton, *Folding dynamics and mechanism of β-hairpin formation*, Nature 390 (2003), pp. 196–199.
[32] K. Klimov and D. Thirumalai, *Mechanisms and kinetics of β-hairpin formation*, Proc. Natl. Acad. Sci. USA 97 (2000), pp. 2544–2549.
[33] B. Zagrovic, E.J. Sorin, and V.S. Pande, *β-hairpin folding simulations in atomistic detail using an implicit solvent model*, J. Mol. Biol. 313 (2001), pp. 151–169.
[34] A.E. García and K.Y. Sanbonmatsu, *Exploring the energy landscape of a β-hairpin in explicit solvent*, Proteins 42 (2001), pp. 345–354.
[35] P.G. Bolhuis, *Transition-path sampling of β-hairpin folding*, Proc. Natl. Acad. Sci. USA 100 (2003), pp. 12129–12134.
[36] N. Kobayashi, S. Honda, H. Yoshii, and E. Munekata, *Role of side chains in the cooperative β-hairpin folding of the short C-terminal fragment derived from streptococcal protein G*, Biochemistry 39 (2000), pp. 6564–6571.
[37] R.H. Zhou, B.J. Berne, and R. Germain, *The free energy landscape for β hairpin folding in explicit water*, Proc. Natl. Acad. Sci. USA 98 (2001), pp. 14931–14936.
[38] J. Tsai and M. Levitt, *Evidence of turn and salt bridge contributions to β-hairpin stability: MD simulations of C-terminal fragment from the B1 domain of protein G*, Biophys. Chem. 101 (2002), pp. 187–201.
[39] Y.Q. Zhou and A. Linhananta, *Role of hydrophilic and hydrophobic contacts in folding of the second β-hairpin fragment of protein G: Molecular dynamics simulation studies of an all-atom model*, Proteins 47 (2002), pp. 154–162.
[40] Xu. Wu, S. Wang, and B.R. Brooks, *Direct observation of the folding and unfolding of a β-hairpin in explicit water through computer simulation*, J. Am. Chem. Soc. 124 (2002), pp. 5282–5283.
[41] G.H. Wei, N. Mousseau, and P. Derreumaux, *Complex folding pathways in a simple β-hairpin*, Proteins 56 (2004), pp. 464–474.
[42] K.A. Olsen, R.M. Fesinmeyer, J.M. Stewart, and N.H. Andersen, *Hairpin folding rates reflect mutations within and remote from the turn region*, Proc. Natl. Acad. Sci. USA 102 (2005), pp. 15483–15487.
[43] P.H. Nguyen, Y.G. Mu, and G. Stock, *Structure and energy landscape of a photoswithchable peptide: A replica exchange molecular dynamics study*, Proteins 60 (2005), pp. 485–494.
[44] G. Bussi, F.L. Gervasio, A. Laio, and M. Parrinello, *Free-energy landscape for beta hairpin folding from combined parallel tempering and metadynamics*, J. Am. Chem. Soc. 128 (2006), pp. 13435–13441.
[45] S. Patel, P. Sista, P.V. Balaji, and Y.U. Sasidhar, *β-hairpins with native-like and non-native hydrogen bonding patterns could form during the refolding of staphylococcal nuclease*, J. Mol. Graph. Model 25 (2006), pp. 103–115.
[46] D.G. Du, M. Tucker, and F. Gai, *Understanding the mechanism of beta-hairpin folding via phi-value analysis*, Biochemistry 45 (2006), pp. 2668–2678.
[47] J. Zhang, M. Qin, and W. Wang, *Folding mechanism of β-hairpins studied by replica exchange molecular simulations*, Proteins 62 (2006), pp. 672–685.
[48] A.M. J. J. Bonvin and W.F. van Gunsteren, *β-hairpin stability and folding: Molecular dynamics studies of the first β-hairpin of tendamistat*, J. Mol. Biol. 296 (2000), pp. 255–268.
[49] D.G. Du, Y.J. Zhu, C.Y. Huang, and F. Gai, *Understanding the key factors that control the rate of β-hairpin folding*, Proc. Natl. Acad. Sci. USA 101 (2004), pp. 15915–15920.
[50] H.J.C. Berendsen, D. van der Spoel, and R. van Drunen, *A message-passing parallel molecular dynamics implementation*, Comp. Phys. Comm. 91 (1995), pp. 43–56.
[51] W.L. Jorgensen, D.S. Maxwell, and J. Tirado-Rives, *Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids*, J. Am. Chem. Soc. 118 (1996), pp. 11225–11236.
[52] B. Hess, H. Bekker, H.J.C. Berendsen, and J.G.E.M. Fraaije, *LINCS: A linear constraint solver for molecular simulations*, J. Comp. Chem. 18 (1997), pp. 1463–1472.
[53] R.V. Pappu, R.K. Hart, and J. W. Ponder, *Analysis and application of potential energy smoothing and search methods for global optimization*, J. Phys. Chem. B 102 (1998), pp. 9725–9742.
[54] W.X. Xu, J. Wang, and W. Wang, *Folding behavior of chaperonin-mediated substrate protein*, Proteins 61 (2005), pp. 777–794.
[55] E. Gallicchio, M. Andrec, A.K. Felts, and R.M. Levy, *Temperature weighted histogram analysis method, replica exchange, and transition paths*, J. Phys. Chem. B 109 (2005), pp. 6722–6731.
[56] G. Zuo, J. Wang, and W. Wang, *Folding with downhill behavior and low cooperativity of proteins*, Proteins 63 (2006), pp. 165–173.
[57] J.D. Chodera, W.C. Swope, J.W. Pitera, C. Seok, and K.A. Dill, *Use of the weighted histogram analysis method for the analysis of simulated and parallel tempering simulations*, J. Chem. Theory. Comput. 3 (2007), pp. 26–41.
[58] A.M. Ferrenberg and R.H. Swendsen, *Optimized Monte Carlo data analysis*, Phys. Rev. Lett. 63 (1989), pp. 1195–1198.
[59] Y.G. Mu, L. Nordenskiold, and J.P. Tam, *Folding, misfolding, and amyloid protofibril formation of WW domain FBP28*, Biophys. J. 90 (2006), pp. 3983–3992.
[60] Y.G. Mu, P.H. Nguyen, and G. Stock, *Energy landscape of a small peptide revealed by dihedral angle principal component analysis*, Proteins 58 (2005), pp. 45–52.
[61] C.D. Snow, L. Qiu, D. Du, F. Gai, S.J. Hagen, and V.S. Pande, *Trp zipper folding kinetics by molecular dynamics and temperature-jump spectroscopy*, Proc. Natl. Acad. Sci. USA 101 (2004), pp. 4077–4082.
[62] A. Baumketner and J.E. Shea, *The thermodynamics of folding of a β-hairpin peptide probed through replica exchange molecular dynamics simulations*, Theor. Chem. Acc. 116 (2006), pp. 262–273.
[63] K.Y. Sanbonmatsu and A.E. García, *Structure of Met-Enkephalin in explicit aqueous solution using replica exchange molecular dynamics*, Proteins 46 (2002), pp. 225–234.
[64] R. Koradi, M. Billeter, and K. Wüthrich, *MOLMOL: A program for display and analysis of macromolecular structures*, J. Mol. Graph. 14 (1996), pp. 51–55.
[65] N.H. Andersen, K.A. Olsen, R.M. Fesinmeyer, X. Tan, F.M. Hudson, L.A. Eidenschink, and S.R. Farazi, *Minimization and optimization of designed β-hairpin folds. ββ multiple turns in proteins*, J. Am. Chem. Soc. 128 (2006), pp. 6101–6110.
[66] S. Kumar and R. Nussinov, *Salt bridge stability in monomeric proteins*, J. Mol. Biol. 293 (1999), pp. 1241–1255.
[67] A.S. Thomas and A.H. Elcock, *Molecular simulations suggest protein salt bridges are uniquely suited to life at high temperatures*, J. Am. Chem. Soc. 126 (2004), pp. 2208–2214.
[68] W.Y. Yang, J.W. Pitera, W.C. Swope, and M. Gruebele, *Heterogeneous folding of the trpzip hairpin: Full atom simulation and experiment*, J. Mol. Biol. 336 (2004), pp. 241–251.
[69] S. Deechongkit, H. Nguyen, E.T. Powers, P.E. Dawson, M. Gruebele, and J.W. Kelly, *Context-dependent contributions of backbone hydrogen bonding to β-sheet folding energetics*, Nature 430 (2004), pp. 101–105.
[70] E.G. Huntchinson and J.M. Thornton, *A revised set of potentials for β-turn formation in proteins*, Protein Sci. 3 (1994), pp. 2207–2216.

[71] K. Guruprasad, M.S. Prasad, and G.R. Kumar, *Analysis of $\gamma\beta$, $\beta\gamma$, $\gamma\gamma$, $\beta\beta$ multiple turns in proteins*, J. Pept. Res. 56 (2000), pp. 250–263.

[72] A.K. Felts, E. Gallicchio, A. Wallqvist, and R.M. Levy, *Distinguishing native conformations of proteins from decoys with an effective free energy estimator based on the OPLS all-atom force field and the surface generalized Born solvent model*, Proteins 48 (2002), pp. 404–422.

[73] J.P. Ulmschneider and W.L. Jorgensen, *Polypeptide folding using Monte Carlo sampling, concerted rotation, and continuum solvation*, J. Am. Chem. Soc. 126 (2004), pp. 1849–1857.

[74] R. Zhou and B.J. Berne, *Can a continuum solvation model reproduce the free energy landscape of a β-hairpin folding in water?* Proc. Natl. Acad. Sci. USA 99 (2002), pp. 12777–12782.

[75] H.Y. Liu and X. Zou, *Electrostatics of ligand binding: Parametrization of the generalized Born model and comparison with the Poisson-Boltzmann approach*, J. Phys. Chem. B 110 (2006), pp. 9304–9313.

[76] J. Chen, W. Im, and C.L. Brooks, *Balancing solvation and intramolecular interactions: Toward a consistent generalized Born force field*, J. Am. Chem. Soc. 128 (2006), pp. 3728–3736.

[77] S.H. Lin, P.K. Maiti, and W.A. Goddart, III, *Dynamics and thermodynamics of water in PAMAM dendrimers at subnanosecond time scales*, J. Phys. Chem. B 109 (2005), pp. 8663–8672.

[78] S. Woutersen, Y.G. Mu, G. Stock, and P. Hamm, *Subpicosecond conformational dynamics of small peptides probed by two-dimensional vibrational spectroscopy*, Proc. Natl. Acad. Sci. USA 98 (2001), pp. 11254–11258.

[79] D. Roe, A. Okur, L. Wickstrom, V. Hornak, and C. Simmerling, *Secondary structure bias in generalized Born solvent models: Comparison of conformational ensembles and free energy of solvent polarization from explicit and implicit solvation*, J. Phys. Chem. B. 111 (2007), pp. 1846–1857.

[80] A. Onufriev, D. Bashford, and D.A. Case, *Exploring protein native states and large-scale conformational changes with a modified generalized Born model*, Proteins 55, pp. 383–394.

[81] M. Stork and P. Tavan, *Electrostatics of proteins in dielectric solvent continua. I. Newton's third law marries qE forces*, J. Chem. Phys. 126 (2007), pp. 165105–165115.